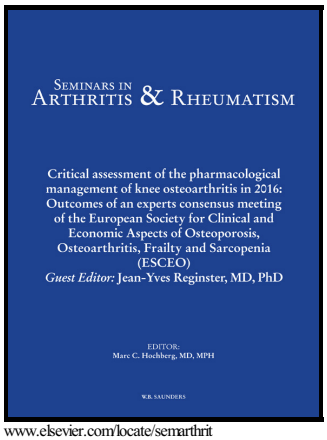

www.elsevier.com/locate/semarthrit

Equivalence in the health assessment questionnaire (HAQ) across socio-demographic determinants: analyses within QUEST-RA

Monika Hifinger, Sam Norton, Sofia Ramiro, Polina Putrik, Tuulikki Sokka-Isler, Annelies Boonen

PII: S0049-0172(17)30173-7
DOI: http://dx.doi.org/10.1016/j.semarthrit.2017.08.003
Reference: YSARH51225

To appear in: *Seminars in Arthritis and Rheumatism*

Cite this article as: Monika Hifinger, Sam Norton, Sofia Ramiro, Polina Putrik, Tuulikki Sokka-Isler and Annelies Boonen, Equivalence in the health assessment questionnaire (HAQ) across socio-demographic determinants: analyses within QUEST-RA, *Seminars in Arthritis and Rheumatism,* http://dx.doi.org/10.1016/j.semarthrit.2017.08.003

Equivalence in the health assessment questionnaire (HAQ) across socio-demographic determinants: analyses within QUEST-RA

Monika. Hifinger[* 1, 2]; Sam Norton[3, 4]; Sofia Ramiro [5]; Polina Putrik[1, 2]; Tuulikki Sokka-Isler[6]; Annelies Boonen[1, 2]

1. CAPHRI Research Institute, Maastricht University, Maastricht, the Netherlands
2. Department of Rheumatology, Maastricht University Medical Center (MUMC), Maastricht, the Netherlands
3. Department of Psychology, King´s College, London, UK
4. Academic Rheumatology, King´s College, London, UK
5. Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands
6. Department of Rheumatology, Jyväskylä Central Hospital, Jyväskylä, Finland

***Corresponding author, email: monikahifinger@gmx.de**

Address:

Maastricht University Medical Centre

Department of Internal Medicine

Division of Rheumatology

Postbus 5800

6202 AZ Maastricht

+31 433 87 7009

Abstract:

Objectives: To investigate potential bias in scores of the Health Assessment Questionnaire (HAQ) related to socio-demographic (SD) background of patients with rheumatoid arthritis (RA)

Methods:

Data from the Quantitative Standard Monitoring of Rheumatoid Arthritis study (QUEST-RA), comprising 9,022 patients were analysed. Physical function was assessed through 30 items of four HAQ versions: the HAQ-Disability scale, HAQ-II, modified HAQ and multi-dimensional HAQ (MD-HAQ). DIF was investigated using item response theory models implemented in a latent variable modelling framework. Models were equivalent to ordinal logistic regression models with HAQ score (item level) as outcome, the latent trait 'physical function' and individual SD factors (age, gender, education, and employment status) as predictors. Next, scores of composite HAQs were adjusted for DIF. To assess the impact of DIF on associations between SD factors and HAQs, multilevel mixed-effect linear regression models with individuals nested in country were estimated with DIF-adjusted or unadjusted HAQ as outcome.

Results:
Relevant DIF (OR>1.1 or <0.90) was found in several HAQ items primarily for age, gender and work-status. Adjustment of composite HAQs for DIF resulted in small increases (Δ0.02-0.07); MD-HAQ best compensated for bias related to SD factors (Δ0.02). In regressions, all SD-factors remained significantly related to DIF-adjusted HAQs, with differences in coefficients largest for gender (Δ0.02-0.07) but overall negligible.

Conclusions:
SD factors produce response bias in individual HAQ items but have little impact on composite HAQs. When interpreting HAQ across SD factors, MD-HAQ is preferred, but caution remains when comparing function across gender.

Key messages

- Socio-demographic factors cause relevant response bias in individual items of the health assessment questionnaire (HAQ)
- Item bias has minor effect on composite HAQs indicating its overall accuracy across socio-demographic groups
- Notwithstanding, interpretation of HAQ outcomes across gender requires some caution
- Use of Multidimensional HAQ (MD-HAQ) suggested when comparing physical function across socio-demographic groups

Keywords: Rheumatoid arthritis, Outcome Research, Patient social context, Patient-reported outcome

Declaration of Interest

Data collection was supported by the Academy of Finland and Abbott. The ancillary study described in this manuscript was done without any type of funding.
MH contributed during an unpaid extended maternity leave (2013-2016) agreed with Hexal AG, Germany. Related to the topic of this study, all other co-authors have no disclosures to declare.

Ethical approval information

The protocol for original QUEST-RA study was reviewed and approved by all local institutional review boards or ethics committees of participating countries.

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory disease associated with varying impairments in daily tasks and activities [1]. In the traditional biomedical paradigm, severity and impact of rheumatic diseases have been mainly assessed using objective measures such as erythrocyte sedimentation rate (ESR), painful and swollen joint counts and radiographic damage. To better understand the impact of the disease on the life of patients, patient reported outcomes (PRO) have been developed. The Health Assessment Questionnaire (HAQ) was among the first patient-reported disability measures validated for use in RA and has become the dominant instrument for assessment of physical function of RA patients in clinical practice and research worldwide [2-4]. HAQ is a predictor of several outcomes including work participation and mortality [5, 6]. Numerous studies provided evidence that variation in HAQ can be explained by (fluctuations in) disease activity as well as by (more permanent) radiographic damage [7, 8].

Today a number of validated versions of the HAQ exist. For several decades the original (Stanford HAQ) often also referred to as *HAQ Disability Index (HAQ-DI)* [9, 10] has been considered the gold standard to measure physical function in RA. However, with 8 domains and overall 41 items, the length of the questionnaire makes its use in clinical practice rather difficult, leading to the development of additional versions of the HAQ intended to be easier to score with better or equal psychometric properties. The *modified HAQ (MHAQ)* [11] is a HAQ version that has a substantially reduced number of items (8 compared to 41) and a simplified scoring system. Another simplified version of the HAQ, called *multi-dimensional (MD-HAQ)* [2] consisting of 10 items has been designed to better detect improvements in function at the lower end of the scale as compared to the MHAQ and incorporates additional items relating to psychological functioning. Addressing this floor effect has become particularly important with increasing efforts to achieve early diagnosis and the paradigm shift towards early treatment. Another version of the HAQ, HAQ II [12] has also been implemented in an attempt to correct the floor effects seen with earlier modifications of the HAQ using a different methodological development approach compared to MD-HAQ. HAQ-II consists of 10 items and is now among the most frequently used HAQ versions [13].

There is increasing awareness that socio-demographic (SD) factors can have an important impact on RA outcomes [14-19]. The magnitude of health disparities related to SD background of patients has been the focus of public policy interest as some of them might be avoidable [20]. With regard to HAQ, a number of studies in RA found relevant differences in HAQ related to education [16] or gender [14]. However, little research has determined the extent to which these observed differences relate to true differences in physical function between SD groups, or are simply caused by biased item responses between SD groups leading to systematically biased interpretation of findings.

A common technique to investigate response bias across items is the analysis of differential item functioning (DIF). DIF occurs when an individual item included in a PRO performs differently for one subgroup (e.g. female) than it does for another (e.g. male) at the same level of the construct it is intended to measure (e.g. physical function). That is to say, the item may be confounded by gender since a person's response to the item may not solely depend on the underlying level of disability.

Despite the widespread use of the HAQ, relatively few studies have investigated DIF for the HAQ items [1, 21-23] and no systematic attention has been given to DIF across different SD groups. Insight into the presence and magnitude of DIF is relevant when interpreting differences in HAQ outcomes between SD determinants and investigating inequalities or

inequities in RA care. We hypothesize that bias will be present across several HAQ items. More specifically, we expect that for the same level of physical function, women, elderly, lower educated and employed persons may experience more difficulty with more strenuous activities affecting mainly the lower limbs. As the HAQ-DI contains the fewest number of items addressing strenuous (lower limb) activities, we expect the HAQ-DI to be least sensitive to the influence of DIF. However, we have no specific expectations about the magnitude of the effect for the total HAQ score, nor about the extent to which DIF can explain the influence of the patient´s SD background on HAQ. In the present study we therefore aimed at investigating potential bias among RA patients in the assessment of physical function across four different versions of the HAQ using DIF. For the analysis, data from QUEST-RA were used to understand presence, magnitude and relevance of measurement equivalence across a number of SD determinants.

The Quantitative Standard Monitoring of Rheumatoid Arthritis study (QUEST-RA) [24] is a multi-centre database that collected SD data as well as extensive data on physical function from more than 10,000 RA patients - a unique setting to undertake the present study.

Materials and Method

Patients and assessments

Baseline data from the international, multi-centre observational study QUEST-RA, were collected between the years 2005 and 2012. The dataset included information from 10,150 RA patients from 34 countries worldwide. The protocol was approved by all local institutional review boards or ethics committees. Patients were eligible to participate in the study if they were at least 18 years of age and fulfilled the 1987 American College of Rheumatology classification criteria for RA. Written informed consent was obtained from all subjects before enrolment.

Following preliminary analysis, data from 4 countries (Egypt, Morocco, Serbia and Russia) were excluded from the present analysis. DIF was observed on the country level to the extent that total scores from these countries were potentially non-comparable with other countries. Since DIF at the country level would interfere with results found for SD factors, they have been excluded [25].

Data collection

Patients completed a self-reported questionnaire, comprising SD background and clinical disease characteristics (PROs), 0-10 visual analogue scales (VAS, 10 worst outcome) for patient global assessment of disease (PatGA), fatigue and pain. Furthermore, patients underwent a clinical examination to assess tender-and swollen joint count (TJC, SJC), and had a laboratory test assessing ESR, C-reactive protein (CRP) and rheumatoid factor (RF). Data on medication intake, disease duration and comorbidities were collected by attending physicians [24, 26, 27]. From the available comorbidities, the Rheumatic Disease Comorbidity Index (RDCI) was calculated [3, 28]. The 28-joint count disease activity score (DAS28) based on 28-swollen and tender joint counts, ESR and PatGA [29] was computed.

Physical function assessment

Four frequently applied and validated versions of the HAQ were used to develop the item pool for the patient survey - the original HAQ-DI as well as three revised versions of the HAQ including the MHAQ, MD-HAQ and HAQ-II. For each

individual HAQ item, the patient could select among four response categories (0 = without any difficulty to, 3 = unable to do). Overall, 30 individual HAQ items were tested in the survey (table 2). Although aids and devices were assessed, these were excluded from the present analyses since they focused on potential bias in the difficulty ratings of the 30 activity items. On the same line, the four items relating to psychological functioning from the MD-HAQ were not used as they target a different dimension than physical functioning.

Of note, the HAQ-DI comprises 20 specific items of physical function and groups them into 8 categories (dressing and grooming, arising, eating, walking, hygiene, reach and activities). The highest scores (scale 0-3) per category are used to compute the average HAQ score across the 8 individual items. The revised versions of the HAQ (HAQ-II, MHAQ, MD-HAQ) reduced the number of items to 10, 8 and 10 items, respectively (table 2). Compared to the HAQ-DI, some items were replaced by new items in HAQ-II, MHAQ, MD-HAQ.

In QUEST-RA, the patient questionnaire (including the individual HAQ items) was translated into each different language by local rheumatologists and translated back into English by professional translators (exceptions: Danish, Finnish and Swedish translation, they were piloted in clinics). Existing official translations of official patient reported outcomes were used whenever available. More details on the translation process can be found elsewhere [24].

Demographic and socio-economic factors explored for differential item functioning

SD determinants of interest available in QUEST-RA included age (in years), gender, work status and level of education [27, 30, 31]. Work status (working yes or no) was classified for further analyses into full- and part-time workers, students and homemakers as “working”, whereas patients that were retired or disabled were considered “not working”. Educational level was assed using the number of years of formal school education.

Statistical Analysis

Item response theory (IRT) models representing the association between an individuals’ observed response to an item and their (unobservable) score on an underlying latent variable (e.g. their true level of disability) were estimated using *Mplus* 7.1 [32]. Specifically, since the HAQ has an ordinal response scale, the estimated models were full information maximum likelihood graded response models [33]. In simple terms, the baseline model, prior to assessing DIF, involves simultaneously estimating an ordinal logistic regression model for each item regressed onto the latent variable. The model thresholds relate to the item *difficulties* and the slope the item *discrimination*. Preliminary analysis examining the fit of the baseline IRT model (not shown) confirmed that all 30 HAQ items formed a unidimensional scale without the need for removing any items.

The overall analysis was conducted in three steps: 1) identifying DIF for individual items, 2) adjusting HAQ version scale scores to remove the confounding effects of DIF, and 3) examining the impact of DIF difference in interpretation for models regressing unadjusted versus DIF-adjusted HAQ scores on SD determinants.

In the first step, DIF by each factors was assessed using the multiple-indicator-multiple-cause (MIMIC) model method [34]. This extends the baseline model to regress item responses onto an additional variable (e.g. age, gender, education or work status) in addition to the latent disability variable. Since disability is controlled for, the odds ratio (OR) relating to the SD factor reflects the direction and magnitude of DIF. For example, an OR of two for female gender indicates that women are twice as likely to report greater difficulty with the activity (e.g. driving) compared to males with the same

disability level. To facilitate interpretation, ORs for age and education were expressed as standardized ORs. In total, 120 individual models were estimated – one for each HAQ item and SD factor pairing (i.e. 30 items times 4 SD variables). A Benjamini-Hochberg false discovery rate correction for multiple-testing [35] was applied to determine statistically significant DIF. DIF was considered clinically important when ORs differed from 1 by more than 10% (OR>1.1 or <0.90), a cut-off frequently used in studies investigating DIF with ordinal logistic regression approaches [36, 37].

In the second step, scores for the four HAQ versions were calculated by estimating a separate IRT models including only the subset of items included in each of the four HAQ versions considered. Using the MIMIC method [34], item responses were regressed on a latent disability variable and whichever SD factors demonstrated significant and clinically important DIF for the item in the first step. The HAQ-DI has a unique scoring system where, of the 20 items, only those with the highest ratings across 8 domains are scored, with further weighting for devices and aids. Since DIF is modelled at the items rather than domain level it is not possible to apply the traditional scoring algorithm. Instead, all 20 items included in the HAQ-DI were used in the IRT MIMIC model. Resulting scores are not directly comparable (i.e. are lower) to those from the traditional HAQ-DI but remove systematic differences that would confound estimates in the following step. Moreover, the scoring of the revised versions of the HAQ avoids this issue, and thus the adjusted scores are more directly comparable when interpreting influence of DIF in our study.

In the final step of the analysis, regression models assessed the degree to which bias at the item level biased the estimates of the association between SD factors and the total scores for each version of the HAQ. The purpose of this step is to demonstrate the impact of DIF using the HAQ in a common analysis and as such determining the degree of bias and demonstrate the validity of other analyses. Separate multilevel linear regression models were estimated for each of the HAQ versions with unadjusted (Model 1, M_1) and DIF adjusted HAQ (Model 2, M_2) scores as outcome. A random effect was estimated to account for the clustering of patients within countries. The four SD factors assessed for DIF were included as predictors in addition to DAS28 and RDCI. Bias was assessed by comparing the difference in regression coefficients for the SD factors between M_1 and M_2. Further assessment of bias was provided by comparing goodness-of-fit using the variance explained by the fixed factors (marginal R-squared) [38]. Relative differences in parameter estimates of ≥ 10% were considered as indicating relevant bias [39]. Regression analyses were estimated in Stata 12 [40].

Results

In total, 9,022 patients from 30 countries were included in the final analyses. Mean age of patients was 55.6 years (SD 13.6), 80.8% were female. On average, patients attended school for 11.0 years (SD 4.1), 54.2% of patients reported to be working (paid in full or part time or unpaid as student or homemaker). Mean composite HAQ scores (mean across items) were 0.7 (SD 0.7) for HAQ-DI, 0.6 (SD 0.6) for MHAQ, 0.8 (SD 0.6) for MD-HAQ and 1.0 (SD 0.7) for HAQ-II. Further descriptive statistics (overall and per country) can be extracted from table 1.

For each of the SD determinants, important and significant DIF (false discovery rate (FDR) p<0.05 and change of standardized OR ≥ 1.1 or ≤0.90) was seen in univariable analysis for several HAQ items (table 2). Age was related to important and significant DIF for 14 of 30 items. Controlling for overall disability, older patients were less likely to indicate difficulty in performing tasks involving hand function (e.g. cut meat, lift cup, open jar, turn faucet) and transfers (e.g. get in/out of bed or on/off the toilet). However, controlling for overall disability, older patients were more likely to indicate

difficulty for physically demanding activities, especially when lower limb function is involved (e.g. run or walk 2 miles, climb stairs, do sports or move/lift heavy objects). Of note, these activities were often part of the revised HAQ versions.
For employment status important and significant DIF was seen in 19 of 30 items. Controlling for overall disability, patients who were in paid or unpaid employment were more likely to report difficulties with tasks involving the hands (e.g. cut meat, open jar) but less likely to report difficulty with more strenuous activities involving lower limb function (e.g. running or walking 2 miles) in univariable analyses.
For gender, important and significant DIF was observed in 23 of 30 HAQ items. Compared to males with the same overall disability, females reported systematically less difficulties for items related to dressing and grooming as well arising, whereas they reported more difficulties for items that require hand strength or are physically more demanding. The highest gender related DIF was found for driving a car with females experiencing more than twice the likelihood of reporting greater difficulty compared to males despite same overall disability level (OR=2.26).
For education, 10 of 30 items showed significant and important DIF but no clear trend could be observed.
To assess the impact of DIF on overall scores on HAQ DIF adjusted HAQ scores were estimated for each version. After adjusting, all mean HAQ scores increased (Δ +0.02 to +0.07, dependent on HAQ version). The smallest increase was observed MD-HAQ (+0.02), which best compensated for the item bias resulting from SD background (table 3).
When comparing crude and DIF-adjusted HAQ scores in multilevel regression models, coefficients for gender increased for all HAQ versions (Δβ: +0.02 to +0.07 corresponding to relative increases +10 to 70%). This increase indicates that effect of DIF between males and females is underestimated in particular for HAQ-DI and MD-HAQ. Coefficients for education, age, DAS28, RDCI as well as country remained largely unchanged (table 3). Overall model fit consistently increased by 2 to 5% in relative terms after adjustment of HAQ for DIF indicating that some systematic measurement error was removed from the total scores (table 3).

Discussion

The present study revealed that for RA patients, the SD background confounded self-reported difficulties in physical function of various individual HAQ items across a range of versions of the HAQ. After adjusting for DIF, all composite HAQ scores increased but overall showed good measurement accuracy across different SD groups. The smallest DIF related changes were observed for MD-HAQ suggesting its use when investigating physical function across different SD groups. Regression analyses revealed that effect of DIF mean differences between males and females were generally underestimated and require some caution when comparing HAQ outcomes between these groups.

Gender DIF has been detected for a number of health outcomes (e.g. pain, fatigue) with women being more likely to report symptoms of distress and thus score worse on experience-based outcomes [41]. Studies specifically investigating DIF related to physical function confirmed the present findings of gender DIF in various individual HAQ items [21, 22]. In line with the DIF patterns observed by Rose *et al* for several items of the HAQ-DI, men did worse with dressing or grooming, women with grip, reach or strength [22]. However, when interpreting these findings, it cannot be ignored that the difference in physical constitution between women and men may also play a role in the experience of physical

abilities. For example, men are usually taller than women possibly contributing to the finding that men experience reaching objects less difficult irrespective of their disease-related functional limitations. Similarly, men are usually stronger possibly contributing to women experiencing more difficulty with lifting or moving heavy objects, cutting meat, opening jar or turning faucets.

When comparing results across age, we are the first to show that in RA higher age was overall associated with higher level of difficulties for many physical demanding activities e.g. walking, climbing stairs, sports, move/lift heavy objects. In an osteoarthritis index [42], DIF was reported for items measuring physical function, with older patients reporting more limitations in climbing stairs than they would have for an unbiased item [43]. IRT models do not enable detection of the specific reason for differential item function, but physical factors (e.g. loss of body strength and fitness), may partly explain the variance. Interestingly older patients tended to experience usual activities of daily living – dressing and grooming, arising, eating, walking, hygiene, reach, grip - as less strenuous.

Work status seemed to be a relevant source of DIF. However, DIF found between working and non-working patients should be interpreted cautiously. In QUEST-RA, patients that work were on average more than 10 years younger than patients that do not work. As such it was not surprising that the findings for work status overlap with findings for age with younger patients being more likely to respond similar to patients that work. While it cannot be excluded that age accounted for some of the observed DIF it is useful to note the greater number of items exhibited DIF by work status than by age. Maybe time pressure and other stress factors that are likely found to a larger extent among working patients contributed to this effect. “Unexpectedly, working patients were more likely to report difficulties for tasks involving the hands but less likely to report difficulties for activities involving lower limb function. Possibly, patients at working age rely to a larger extent on hand function for daily tasks (e.g. office employees, students) and thus have greater demands for hand function. The fact that working patients report less difficulties for strenuous activities may partly relate to a selection effect with working patients overall characterized by better overall health and body constitution (“healthy workers”) that allows these patients to experience less difficulties for activities such as queuing, walking or running moderate distances .It is of note, that the groups “not employed” is a mixture of persons that are retired due to age, work disabled or economically unemployed”

Interestingly, only small effects were observed for education as potential source of DIF. Item bias for education was found in several earlier studies. Perkins *et al.* [44] investigated DIF on the Short Form-36 health survey among healthy individual and found that respondents with less education had less difficulty with vigorous activities at all values of the physical functioning scale. Likely, item bias on education is more important for cognitive tests rather than for tests on physical function [20].

Although a number of individual HAQ items showed significant DIF across all SD factors, it seems that the magnitude of effects minimize in the total score of different versions of the HAQ. After adjusting total HAQ scores for DIF, the mean scores increased by < 0.1 for all versions of the HAQ. Of note, although the minimal clinically important difference (MCID) has been developed to understand a clinically relevant difference in physical function for an individual patient, it can be helpful to realize that the MCID for HAQ (measured on a 0-3 scale) is considered between 0.2 and 0.3 [9, 45-47].

Regression analyses revealed that coefficients of clinical confounders and most SD factors remain largely unaffected by the adjustment of composite HAQs for DIF indicating that observed DIF has very limited influence on the association between HAQ and clinical and SD confounders. However, some changes in coefficient were observed for gender indicating that some measurement equivalence cannot be ignored; the clinical relevance of effects however is expected to be small.

Some limitations have to be considered when interpreting study results. First, although a rigorous translation process was followed, linguistic or translational bias cannot be fully excluded potentially influencing the interpretation of HAQ items. However, across countries and languages, the study recruited patients of diverse SD background so that potential linguistic effects in some individual countries are not expected to influence overall results. Second, the database did not allow for analysis of further possible socio-economic factors, e.g. the influence of race, religion or individual economic background due to lack of adequate variables. Following the PROGRESS framework [48], these factors may contribute to different perception towards health. Finally, the analysis investigated only uniform DIF relating to item difficulties. Non-uniform DIF, where differences in item discrimination is also assessed, was not assessed but could be examined in further analyses.

This study is among the first to investigate item bias when assessing physical function across multiple SD factors. QUEST-RA included data on four different versions of the HAQ and thus allowed comparison of different clinically relevant tools to assess physical function. The analysis included almost 9,000 patients from 30 countries and thus provides a unique opportunity to overcome limitations on generalizability of results.

Conclusions:

Although important item bias between different SD groups could be found for individual HAQ items, composite HAQ scores overall measure accurately the level of physical function across different SD groups. Still there is some need for caution when comparing physical function between males and females.

Acknowledgements

The QUEST-RA Investigators:

Argentina: Sergio Toloza, Santiago Aguero, Sergio Orellana Barrera, Soledad Retamozo**,** Hospital San Juan Bautista, Catamarca, Paula Alba, Cruz Lascano, Alejandra Babini, Eduardo Albiero, Hospital of Cordoba, Cordoba, Eduardo Kerzberg, Ramos Mejía Hospital, Buenos Aires, Argentina

Brazil: Geraldo da Rocha Castelar Pinheiro, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Licia Maria Henrique da Mota, Hospital Universitário de Brasília, Ines Guimaraes da Silveira, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, FAC Rocha, Universidade Federal do Ceará, Fortaleza, Ieda Maria Magalhães Laurindo, Universidade Estadual de São Paulo, São Paulo

Canada: Juris Lazovskis, Riverside Professional Center, Sydney, NS

Denmark: Merete Lund Hetland, Lykke Ørnbjerg, Kathrine Grøn, Glostrup Hospital, Glostrup and University of Copenhagen, Copenhagen, Denmark, Kim Hørslev-Petersen, King Christian the Xth Hospital, Gråsten, Troels Mørk Hansen, Lene Surland Knudsen, Copenhagen Univ Hospital at Herlev, Herlev

Egypt: Hisham Hamoud, Mohamad Sobhy, Ahmad Fahmy, Mohamad Magdy, Hany Aly, Hatem Saeid, Ahmad Nagm, Al-Azhar University, Cairo, Nihal A Fathi, Essam A Abda, Assiut University Hospital, Zahraa I Selim, Sohage University Hospital, Sohag

Estonia: Raili Müller, Reet Kuuse, Marika Tammaru, Riina Kallikorm, Tartu University Hospital, Tartu, Tony Peets, Kati Otsa, Karin Laas, East-Tallinn Central Hospital, Tallinn, Ivo Valter, Center for Clinical and Basic Research, Tallinn

Finland: Tuulikki Sokka-Isler, Jyväskylä Central Hospital, Jyväskylä, Kai Immonen, Jukka Lähteenmäki, North Karelia Central Hospital, Joensuu, Timo Yli-Kerttula, Päivi Ekman, Satakunta Central Hospital, Rauma; Markku Kauppi, Lahti Central Hospital, Lahti
France: Laure Gossec, Maxime Dougados, University René Descartes, Hôpital Cochin, Paris, Jean Francis Maillefert, Dijon University Hospital, University of Burgundy, and INSERM U887, Dijon, Bernard Combe, Hôpital Lapeyronie, Montpellier, Jean Sibilia, Hôpital Hautepierre, Strasbourg
Germany: Siegfrid Wassenberg, Evangelisches Fachkrankenhaus, Ratingen, Rieke Alten, Christof Pohl, Schlosspark-Klinik, Berlin, Gerd R Burmester, Bettina Marsmann, Jacqueline Detert, Charite – University Medicine Berlin, Berlin
Greece: Alexandros A. Drosos, Sofia Exarchou, University of Ioannina, Ioannina, H M Moutsopoulos, Afrodite Tsirogianni, School of Medicine, National University of Athens, Athens, Fotini N Skopouli, Maria Mavrommati, Euroclinic Hospital, Athens
Hungary: Pál Géher, Semmelweis University of Medical Sciences, Budapest, Bernadette Rojkovich, Ilona Újfalussy, Polyclinic of the Hospitaller Brothers of St. John of God in Budapest, Budapest
Ireland: Patricia Minnock, Our Lady's Hospice, Dublin, Eithne Murphy, Claire Sheehy, Edel Quirke, Connolly Hospital, Dublin, Joe Devlin, Shafeeq Alraqi, Waterford Regional Hospital, Waterford
India: Amita Aggarwal, Department of Immunology, Lucknow, Sapan C Pandya, Vedanta Institute of medical Sciences, Ahmedabad, Banwari Sharma, Department of Immunology, Jaipur Hospital**,** Jaipur
Italy: Massimiliano Cazzato, Stefano Bombardieri, Santa Chiara Hospital, Pisa, Gianfranco Ferraccioli, Alessia Morelli, Catholic University of Sacred Heart, Rome, Sabrina Paolino, Maurizio Cutolo, University of Genova, Genova, Italy, Fausto Salaffi, Andrea Stancati, University of Ancona, Ancona**,**
Japan: Hisashi Yamanaka, Ayako Nakajima, Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Wataru Fukuda, Department of Rheumatology, Kyoto First Red Cross Hospital, Kyoto, Eisuke Shono, Shono Rheumatism Clinic, Fukuoka
Kenya: G Omondi Oyoo, Kenyatta Hospital, Nairobi
Korea: Shin-Seok Lee, Chonnam National University Medical School, Gwangju; Jung-Yoon Choe, Catholic University of Daegu School of Medicine; Daegu, Eun Bong Lee, Seoul National University College of Medicine, Seoul, Kichul Shin, SMG-SNU Boramae Medical Center, Seoul
Kosovo: Sylejman Rexhepi, Mjellma Rexhepi, Blerta Rexhepi**,** Rheumatology Department, Pristine
Latvia: Daina Andersone, Paula Stradina Clinical University Hospital, Riga
Lithuania: Sigita Stropuviene, State Research Institute for Inovative
Medicine and Vilnius University Medical Faculty, Jolanta Dadoniene, Vilnius University Medical Faculty, Vilnius, Asta Baranauskaite, Kaunas University Hospital, Kaunas
Morocco: Najia Hajjaj-Hassouni, Karima Benbouazza, Fadoua Allali, Rachid Bahiri, Bouchra Amine, El Ayachi Hospital Mohamed Vth Souissi University, Rabat
Netherlands: Johannes WG Jacobs, Suzan MM Verstappen, University Medical Center Utrecht, Utrecht, Margriet Huisman, Femke Bonte-Mineur, Chiara Messidoro, Sint Franciscus Gasthuis, Rotterdam, Hans Rasker, Monique Hoekstra, Medisch Spectrum Twente, Enschede
Norway: Glenn Haugeberg, Hilde Gjelberg, Eirik Wilberg, Sørlandet Hospital, Kristiansand
Poland: Stanislaw Sierakowski, Medical University in Bialystok, Bialystok, Maria Majdan, Medical University of Lublin, Lublin, Wojciech Romanowski, Poznan Rheumatology Center in Srem, Srem, Witold Tlustochowicz, Military Institute of Medicine, Warsaw, Danuta Kapolka, Silesian Hospital for Rheumatology and Rehabilitation in Ustron Slaski, Ustroñ Slaski, Stefan Sadkiewicz, Szpital Wojewodzki im. Jana Biziela, Bydgoszcz, Danuta Zarowny-Wierzbinska, Wojewodzki Zespol Reumatologiczny im. dr Jadwigi Titz-Kosko, Sopot
Romania: Ruxandra Ionescu, Denisa Predeteanu, Spitalul Clinic Sf Maria, Bucharest, Rodica Marieta Chirieac, Codrina Ancuta, Gr.T.Popa University of Medicine and Pharmacy Iasi, Iasi
Russia: Dmitry Karateev, Elena Luchikhina, Institute of Rheumatology of Russian Academy of Medical Sciences, Moscow, Natalia Chichasova, Moscow Medical Academy, Moscow, Vladimir Badokin, Russian Medical Academy of Postgraduate Education, Moscow, Ivan Shirinsky, Polovnikova Oksana, Institute of Clinical Immunology RAMS, Novosibirsk
Serbia: Vlado Skakic, Aleksander Skakic, Jovan Nedovic, Rheumatology Institut, Niska Banja
Spain: Antonio Naranjo, Carlos Rodríguez-Lozano, Hospital de Gran Canaria Dr. Negrin, Las Palmas, Jaime Calvo-Alen, Hospital Sierrallana Ganzo, Torrelavega, Miguel Belmonte, Hospital General de Castellón, Castellón
Sweden: Eva Baecklund, Uppsala University Hospital, Dan Henrohn, Department of Medical Sciences, Uppsala University, Uppsala, Margareth Liveborn, Centrallasarettet, Västerås, Ann-Carin Holmqvist, Hudiksvall Medical Clinic, Hudiksvall

Turkey: Feride Gogus, Gazi University Medical Faculty, Ankara, Recep Tunc, Meram Medical Faculty, Konya, Selda Celic, Cerrahpasa Medic Faculty, Nevsun Inanc, Haner Direskeneli, Marmara University, Istanbul
Taiwan: Hui-Chu Lang, National Yang-Ming University, Taipei, Hsiao-Yi Lin, MD Taipei Veterans General Hospital, Taipei, Ying-Ming Chiu, MD Changhua Christian Hospital, Changhus, Shinn-shing Lee, MD Cheng Hsin Hospital, Taipei
United Arab Emirates: Humeira Badsha, Dubai Bone and Joint Center, Dubai, Ayman Mofti, "Albiraa clinic" in Dubai, Dubai
United Kingdom: Peter Taylor, Catherine McClinton, Charing Cross Hospital, London, Anthony Woolf, Ginny Chorghade, Royal Cornwall Hospital, Truro, Frances Borg, Essex University Southend University Hospital, Westcliff-on-Sea, Essex
United States of America: Martin Bergman, Taylor Hospital, Ridley Park, PA, Jurgen Craig-Muller, CentraCare Clinic, St. Cloud, MN, Ruben A Peredo, Univ of Michigan Health System, Ann Arbor, MI
Study Center: Tuulikki Sokka, Elena Nikiphorou, Hannu Kautiainen, Jyväskylä Central Hospital, Jyväskylä; Medcare Oy, Äänekoski, Finland, Nasim Khan, University of Arkansas for Medical Sciences, Little Rock, AR, USA

References

1. Oude Voshaar MA, Glas CA, ten Klooster PM, Taal E, Wolfe F, van de Laar MA. Crosscultural measurement equivalence of the Health Assessment Questionnaire II. Arthritis Care Res (Hoboken). 2013 Jun; 65(6):1000-1004.
2. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis and rheumatism. 1999 Oct; 42(10):2220-2230.
3. Wolfe F, Michaud K, Li T, Katz RS. Chronic conditions and health problems in rheumatic diseases: comparisons with rheumatoid arthritis, noninflammatory rheumatic disorders, systemic lupus erythematosus, and fibromyalgia. The Journal of rheumatology. 2010 Feb; 37(2):305-315.
4. Wolfe F, Pincus T. Listening to the patient: a practical guide to self-report questionnaires in clinical care. Arthritis and rheumatism. 1999 Sep; 42(9):1797-1808.
5. de Croon EM, Sluiter JK, Nijssen TF, Dijkmans BA, Lankhorst GJ, Frings-Dresen MH. Predictive factors of work disability in rheumatoid arthritis: a systematic literature review. Annals of the rheumatic diseases. 2004 Nov; 63(11):1362-1367.
6. Callahan LF, Pincus T, Huston JW, 3rd, Brooks RH, Nance EP, Jr., Kaye JJ. Measures of activity and damage in rheumatoid arthritis: depiction of changes and prediction of mortality over five years. Arthritis care and research : the official journal of the Arthritis Health Professions Association. 1997 Dec; 10(6):381-394.
7. Welsing PM, van Gestel AM, Swinkels HL, Kiemeney LA, van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. Arthritis and rheumatism. 2001 Sep; 44(9):2009-2017.
8. Kapetanovic MC, Lindqvist E, Nilsson JA, Geborek P, Saxne T, Eberhardt K. Development of functional impairment and disability in rheumatoid arthritis patients followed for 20 years: relation to disease activity, joint damage, and comorbidity. Arthritis Care Res (Hoboken). 2015 Mar; 67(3):340-348.
9. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health and quality of life outcomes. 2003; 1:20.
10. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis and rheumatism. 1980 Feb; 23(2):137-145.
11. Pincus T, Summey JA, Soraci SA, Jr., Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis and rheumatism. 1983 Nov; 26(11):1346-1353.
12. Wolfe F, Michaud K, Pincus T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. Arthritis and rheumatism. 2004 Oct; 50(10):3296-3305.
13. Maska L, Anderson J, Michaud K. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved

Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). Arthritis Care Res (Hoboken). 2011 Nov; 63 Suppl 11:S4-13.
14. Massardo L, Pons-Estel BA, Wojdyla D, Cardiel MH, Galarza-Maldonado CM, Sacnun MP, et al. Early rheumatoid arthritis in Latin America: low socioeconomic status related to high disease activity at baseline. Arthritis Care Res (Hoboken). 2012 Aug; 64(8):1135-1143.
15. Sokka T, Kautiainen H, Pincus T, Toloza S, da Rocha Castelar Pinheiro G, Lazovskis J, et al. Disparities in rheumatoid arthritis disease activity according to gross domestic product in 25 countries in the QUEST-RA database. Annals of the rheumatic diseases. 2009 Nov; 68(11):1666-1672.
16. Pincus T, Callahan LF. Formal education as a marker for increased mortality and morbidity in rheumatoid arthritis. Journal of chronic diseases. 1985; 38(12):973-984.
17. Carmona L, Loza E. Despair on disparities. Annals of the rheumatic diseases. 2009 Nov; 68(11):1657-1658.
18. Putrik P, Ramiro S, Keszei AP, Hmamouchi I, Dougados M, Uhlig T, et al. Lower education and living in countries with lower wealth are associated with higher disease activity in rheumatoid arthritis: results from the multinational COMORA study. Annals of the rheumatic diseases. 2015 Jan 28.
19. Putrik P, Ramiro S, Hifinger M, Keszei AP, Hmamouchi I, Dougados M, et al. In wealthier countries, patients perceive worse impact of the disease although they have lower objectively assessed disease activity: results from the cross-sectional COMORA study. Annals of the rheumatic diseases. 2016 Apr; 75(4):715-720.
20. Teresi JA, Fleishman JA. Differential item functioning and health assessment. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2007; 16 Suppl 1:33-42.
21. ten Klooster PM, Taal E, van de Laar MA. Rasch analysis of the Dutch Health Assessment Questionnaire disability index and the Health Assessment Questionnaire II in patients with rheumatoid arthritis. Arthritis and rheumatism. 2008 Dec 15; 59(12):1721-1728.
22. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). Journal of clinical epidemiology. 2008 Jan; 61(1):17-33.
23. Kucukdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A. Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis and rheumatism. 2004 Feb 15; 51(1):14-19.
24. Sokka T, Kautiainen H, Toloza S, Makinen H, Verstappen SM, Lund Hetland M, et al. QUEST-RA: quantitative clinical assessment of patients with rheumatoid arthritis seen in standard rheumatology care in 15 countries. Annals of the rheumatic diseases. 2007 Nov; 66(11):1491-1496.
25. Norton S, Hifinger M, Ramiro S, Putrik P, Sokka-Isler T, Boonen A. Comparability of the health assessment questionnaire between countries: Psychometric examination of cross-national measurement equivalence Annals of the rheumatic diseases. 2016; 75(2):165.
26. Gron KL, Ornbjerg LM, Hetland ML, Aslam F, Khan NA, Jacobs JW, et al. The association of fatigue, comorbidity burden, disease activity, disability and gross domestic product in patients with rheumatoid arthritis. Results from 34 countries participating in the Quest-RA program. Clinical and experimental rheumatology. 2014 Nov-Dec; 32(6):869-877.
27. Sokka T, Toloza S, Cutolo M, Kautiainen H, Makinen H, Gogus F, et al. Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study. Arthritis research & therapy. 2009; 11(1):R7.
28. England BR, Sayles H, Mikuls TR, Johnson DS, Michaud K. Validation of the rheumatic disease comorbidity index. Arthritis Care Res (Hoboken). 2015 May; 67(6):865-872.
29. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis and rheumatism. 1995 Jan; 38(1):44-48.
30. Jawaheer D, Olsen J, Lahiff M, Forsberg S, Lahteenmaki J, da Silveira IG, et al. Gender, body mass index and rheumatoid arthritis disease activity: results from the QUEST-RA Study. Clinical and experimental rheumatology. 2010 Jul-Aug; 28(4):454-461.

31. Sokka T, Kautiainen H, Pincus T, Verstappen SM, Aggarwal A, Alten R, et al. Work disability remains a major problem in rheumatoid arthritis in the 2000s: data from 32 countries in the QUEST-RA study. Arthritis research & therapy. 2010; 12(2):R42.
32. Muthén, L.K. and Muthén, B.O. (1998-2015). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén 2015.
33. Samejima F. Estimation of latent ability using response pattern of graded scores 1 ETS Research Bulletin Series. 1968; 1968(1):i-169.
34. Yu YF, Yu AP, Ahn J. Investigating differential item functioning by chronic diseases in the SF-36 health survey: a latent trait analysis using MIMIC models. Med Care. 2007 Sep; 45(9):851-859.
35. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57(1):289-300.
36. Pollard B, Johnston M, Dixon D. Exploring differential item functioning in the SF-36 by demographic, clinical, psychological and social factors in an osteoarthritis population. BMC musculoskeletal disorders. 2013 Dec 11; 14:346.
37. Crane PK, Gibbons LE, Jolley L, van Belle G, Selleri R, Dalmonte E, et al. Differential item functioning related to education and age in the Italian version of the Mini-mental State Examination. Int Psychogeriatr. 2006 Sep; 18(3):505-515.
38. Snijders TAB, Bosker RJ. Multilevel analysis : an introduction to basic and advanced multilevel modeling. London: SAGE, 1999.
39. Cohen J. A power primer. Psychol Bull. 1992 Jul; 112(1):155-159.
40. StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.
41. McHorney CA, Fleishman JA. Assessing and understanding measurement equivalence in health outcome measures. Issues for further quantitative and qualitative inquiry. Med Care. 2006 Nov; 44(11 Suppl 3):S205-210.
42. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. The Journal of rheumatology. 1988 Dec; 15(12):1833-1840.
43. Pollard B, Johnston M, Dixon D. Exploring differential item functioning in the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). BMC musculoskeletal disorders. 2012; 13:265.
44. Fleishman JA, Lawrence WF. Demographic variation in SF-12 scores: true differences or differential item functioning? Med Care. 2003 Jul; 41(7 Suppl):III75-III86.
45. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. The Journal of rheumatology. 2007 Feb; 34(2):280-289.
46. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. The Journal of rheumatology. 1993 Mar; 20(3):557-560.
47. Wolfe F, Michaud K, Strand V. Expanding the definition of clinical differences: from minimally clinically important differences to really important differences. Analyses in 8931 patients with rheumatoid arthritis. The Journal of rheumatology. 2005 Apr; 32(4):583-589.
48. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. Journal of clinical epidemiology. 2014 Jan; 67(1):56-64.

Table 1 - Patient characteristics and scores on the 4 different HAQ versions for the total group and per country

Country	No of patients	Age	Gender	Education	Work status	Mean Health Assessment Questionnaire (HAQ) scores				DAS28	Disease Duration
		In years (mean; (SD))	Female (%)	In years (mean; (SD))	Working* (%)	HAQ-DI** (mean; SD)	MHAQ (mean; (SD)	MD-HAQ (mean; (SD))	HAQ-II (mean; (SD))	(mean; SD)	In years (mean, SD)
Denmark	301	57.8 (13.6)	76.7%	10.7 (3.4)	40.8%	0.5 (0.6)	0.4 (0.5)	0.6 (0.6)	0.8 (0.7)	3.3 (1.5)	9.8 (10.8)
Finland	304	58.5 (12.2)	72.4%	10.4 (3.5)	30.0%	0.5 (0.6)	0.5 (0.5)	0.6 (0.6)	0.9 (0.7)	3.3 (1.4)	11.9 (10.4)
France	388	55.4 (13.5)	78.1%	10.8 (3.6)	46.1%	0.6 (0.5)	0.5 (0.5)	0.7 (0.5)	1.0 (0.6)	3.7 (1.5)	11.2 (9.5)
Germany	224	59.0 (12.7)	83.5%	10.4 (2.6)	39.9%	0.7 (0.6)	0.6 (0.6)	0.8 (0.6)	1.0 (0.7)	4.4 (1.7)	10.6 (9.4)
Ireland	240	56.4 (13.7)	64.3%	11.8 (3.5)	61.0%	0.6 (0.5)	0.5 (0.5)	0.7 (0.6)	0.9 (0.7)	4.1 (1.6)	10.1 (10.5)
Italy	336	61.0 (13.9)	78.2%	8.5 (3.8)	49.7%	0.8 (0.7)	0.7 (0.7)	0.9 (0.7)	1.1 (0.8)	4.5 (1.5)	8.5 (7.8)
Netherlands	317	59.2 (13.7)	66.3%	11.3 (3.4)	53.6%	0.5 (0.5)	0.5 (0.5)	0.7 (0.6)	0.9 (0.7)	3.1 (1.3)	8.2 (7.4)
Poland	642	53.2 (13.7)	86.7%	12.0 (3.4)	35.2%	1.1 (0.7)	0.9 (0.6)	1.1 (0.6)	1.6 (0.7)	5.3 (1.4)	10.2 (8.8)
Spain	302	59.8 (14.5)	73.5%	10.3 (4.9)	52.7%	0.7 (0.7)	0.6 (0.6)	0.8 (0.7)	1.1 (0.8)	3.6 (1.4)	8.1 (6.7)
Sweden	260	59.4 (13.1)	71.8%	10.4 (3.3)	42.1%	0.6 (0.5)	0.5 (0.5)	0.8 (0.5)	1.0 (0.6)	3.8 (1.6)	10.2 (10.1)
UK	191	60.4 (13.5)	73.3%	12.6 (3.1)	38.7%	0.6 (0.6)	0.6 (0.6)	0.8 (0.6)	1.0 (0.7)	3.9 (1.4)	13.2 (11.1)
Turkey	492	52.3 (12.3)	83.2%	7.2 (3.8)	79.2%	0.7 (0.6)	0.5 (0.6)	0.7 (0.6)	1.1 (0.7)	3.9 (1.4)	9.0 (7.9)
USA	401	58.7 (14.0)	74.9%	13.4 (3.1)	59.8%	0.5 (0.5)	0.4 (0.5)	0.7 (0.5)	0.8 (0.7)	3.1 (1.5)	8.6 (9.2)
Argentina	547	53.2 (14.1)	89.1%	8.9 (4.2)	63.9%	0.7 (0.7)	0.6 (0.6)	0.8 (0.7)	1.1 (0.8)	4.4 (1.7)	9.6 (8.6)
Estonia	167	56.1 (13.4)	85.4%	12.5 (2.9)	45.6%	0.9 (0.7)	0.8 (0.6)	1.0 (0.6)	1.2 (0.7)	4.7 (1.5)	8.7 (8.9)
Lithuania	300	54.1 (13.3)	82.9%	12.9 (3.6)	39.2%	1.1 (0.6)	1.0 (0.6)	1.2 (0.6)	1.4 (0.6)	5.5 (1.3)	8.5 (8.4)
Latvia	117	52.6 (12.1)	80.3%	13.0 (3.4)	45.4%	1.2 (0.6)	1.1 (0.6)	1.3 (0.6)	1.4 (0.6)	5.3 (1.5)	11.8 (9.0)
Hungary	153	57.9 (13.6)	87.4%	12.8 (3.2)	19.0%	1.0 (0.6)	0.9 (0.6)	1.1 (0.6)	1.3 (0.7)	5.1 (1.2)	10.6 (9.0)
Greece	299	58.2 (13.6)	75.6%	10.0 (4.9)	57.5%	0.5 (0.6)	0.3 (0.5)	0.5 (0.6)	0.6 (0.7)	3.4 (1.5)	9.8 (7.9)
Canada	100	57.9 (11.5)	78.8%	12.2 (2.6)	33.7%	0.8 (0.6)	0.7 (0.6)	0.9 (0.6)	1.1 (0.7)	4.2 (1.6)	10.1 (10)
UAE	228	45.7 (12.1)	87.6%	14.1 (4.1)	83.7%	0.5 (0.5)	0.4 (0.5)	0.5 (0.5)	0.8 (0.6)	4.2 (1.7)	5.5 (6.1)
Kosovo	100	55.0 (12.0)	84.0%	10.1 (4.4)	52.3%	1.4 (0.5)	1.3 (0.5)	1.5 (0.5)	1.7 (0.5)	6.0 (1.0)	7.4 (5.8)
Brazil	204	51.4 (12.1)	89.6%	9.3 (4.9)	43.0%	0.6 (0.6)	0.6 (0.6)	0.7 (0.6)	0.9 (0.7)	4.2 (1.5)	7.6 (6.9)
Japan	299	59.1 (12.6)	81.2%	12.8 (2.4)	73.7%	0.6 (0.7)	0.5 (0.6)	0.7 (0.7)	0.8 (0.8)	3.8 (1.3)	8.3 (7.2)
India	301	45.5 (10.7)	84.4%	12.0 (4.6)	73.1%	0.6 (0.5)	0.5 (0.5)	0.7 (0.5)	1.1 (0.7)	4.7 (1.4)	4.4 (4.7)
Norway	200	58.0 (13.0)	65.7%	12.1 (3.5)	34.7%	0.5 (0.5)	0.5 (0.4)	0.7 (0.5)	0.7 (0.5)	3.6 (1.4)	11.6 (10.8)

Kenya	388	51.1 (14.6)	92.5%	13.2 (4.0)	69.7%	0.6 (0.6)	0.5 (0.6)	0.7 (0.6)	1.0 (0.7)	4.3 (1.4)	5.2 (6.6)
Romania	322	57.3 (11.9)	87.5%	10.7 (3.7)	16.7%	1.3 (0.7)	1.1 (0.7)	1.3 (0.7)	1.7 (0.7)	5.4 (1.5)	9.3 (7.8)
Korea	603	55.1 (12.2)	86.1%	10.8 (4.0)	84.9%	0.4 (0.5)	0.3 (0.5)	0.4 (0.5)	0.7 (0.6)	3.4 (1.4)	9.1 (7.7)
Taiwan	296	55.1 (12.1)	82.2%	10.9 (4.1)	69.8%	0.2 (0.3)	0.2 (0.3)	0.3 (0.4)	0.5 (0.5)	3.7 (1.5)	8.0 (7.4)
Total	**9,022**	**55.6 (13.6)**	**80.8%**	**11.0 (4.1)**	**54.2%**	**0.7 (0.7)**	**0.6 (0.6)**	**0.8 (0.6)**	**1.0 (0.7)**	**4.1 (1.6)**	**9.1 (8.6)**

*Working includes (full- and part-time employment, unpaid homemaker and students), **HAQ_DI calculated as mean across 20 individual items, HAQ-DI= HAQ-Disability, MHAQ=Modified HAQ, MD-HAQ=MultiDimensional HAQ, DAS28=disease activity score in 28 joints

Table 2: Differential Item Functioning (DIF) across different socio-demographic factors

		Health Assessment Questionnaire HAQ version					Univariable DIF (Odds Ratio (OR), 95% CI)			
	HAQ item, short description [1)]	**HAQ-DI**	**MHAQ**	**MD-HAQ**	**HAQ-II**	**Missing %**	**Age** (per 10 years of age)	**Education** (standardized OR, $SD_{Education}$=4.1 years)	**Gender** (Reference : male)	**Work status** (Reference : not working)
	Dressing & Grooming									
1	dress	✓	✓	✓		0.7	0.91* (0.88,0.95)	1.05 (0.99,1.09)	**0.56* (0.50,0.64)**	1.05 (0.94,1.17)
2	shampoo	✓				1.9	0.96 (0.92,1.01)	1.03 (0.00,1.10)	**0.78* (0.68,0.89)**	1.05 (0.94,1.18)
	Arising									
3	stand up	✓			✓	0.8	0.99 (0.95,1.03)	**1.11* (1.05,1.17)**	**0.77* (0.67,0.88)**	1.05 (0.94,1.17)
4	get in/out of bed	✓	✓	✓		1.3	**0.87* (0.84,0.90)**	1.07* (1.01,1.13)	**0.67* (0.58,0.76)**	**1.38* (1.23,1.54)**
	Eating									
5	cut meat	✓				1.1	**0.87* (0.85,0.92)**	0.98 (0.93,1.03)	**1.27* (1.11,1.45)**	**1.21* (1.08,1.34)**
6	lift cup	✓	✓	✓		1.1	**0.88* (0.84,0.91)**	**1.12* (1.06,1.18)**	0.89 (0.77,1.02)	**1.17* (1.04,1.32)**
7	open milk	✓				1.8	0.96* (0.92,0.99)	1.09* (1.04,1.15)	**1.16* (1.03,1.31)**	**1.16* (1.05,1.27)**
	Walking									
8	walk outdoor	✓	✓	✓	✓	1.1	0.95* (0.91,0.99)	1.05 (1.00,1.12)	**0.69* (0.61,0.80)**	**1.17* (1.04,1.31)**
9	climb 5 step	✓				1.6	0.96 (0.93,1.00)	**0.90* (0.86,0.96)**	0.88 (0.77,1.00)	**1.14* (1.02,1.27)**
	Hygiene									
10	wash body	✓	✓	✓		0.7	0.93* (0.89,0.97)	**1.10* (1.03,1.17)**	**0.54* (0.47,0.62)**	1.02 (0.90,1.15)
11	take bath	✓				5.3	**1.12* (1.07,1.15)**	**1.13* (1.08,1.19)**	1.06 (0.94,1.20)	**0.64* (0.57,0.70)**
12	get on/off toilet	✓			✓	1.3	**0.80* (0.77,0.83**	1.07 (1.00,1.13)	**0.83* (0.73,0.96)**	**1.39* (1.23,1.57)**

								)		
	Reach									
13	reach objects	✓			✓	1.1	1.02 (0.98,1.05)	1.04 (0.99,1.09)	**1.52*** **(1.34,1.71)**	0.99 (0.90,1.09)
14	bend down	✓	✓	✓		1.3	0.96 (0.92,1.00)	0.99 (0.94,1.05)	**0.68*** **(0.60,0.77)**	1.02 (0.91,1.13)
	Grip									
15	open car	✓			✓	1.5	0.91* (0.87,0.95)	0.96 (0.91,1.02)	**1.75*** **(1.51,2.03)**	**1.22*** **(1.09,1.37)**
16	open jars	✓				1.1	**0.88*** **(0.85,0.91)**	1.02 (0.98,1.08)	**1.43*** **(1.34,1.61)**	**1.52*** **(1.38,1.68)**
17	turn faucets	✓	✓	✓		2.0	**0.84*** **(0.83,0.90)**	1.09* (1.04,1.15)	**1.53*** **(1.28,1.75)**	**1.32*** **(1.19,1.47)**
	Activities									
18	run errands	✓				1.7	0.98 (0.94,1.02)	**0.84*** **(0.79,0.89)**	**1.33*** **(1.16,1.54)**	**1.15*** **(1.03,1.29)**
19	get in/out of car	✓	✓	✓		1.3	1.02 (0.98,1.06)	1.00 (0.94,1.05)	**0.77*** **(0.67,0.89)**	0.93 (0.83,1.05)
20	do chores	✓				2.0	0.98 (0.95,1.02)	**1.12*** **(1.05,1.15)**	**1.40*** **(1.24,1.59)**	1.02 (0.93,1.13)
	Item variations[2)]									
21	walk 2 miles			✓		1.8	**1.14*** **(1.10,1.18)**	0.95* (0.90,0.99)	0.91 (0.81,1.02)	**0.78*** **(0.71,0.85)**
22	run 2 miles					4.2	**1.36*** **(1.32,1.41)**	0.92* (0.87,0.97)	**0.81*** **(0.72,0.91)**	**0.61*** **(0.55,0.68)**
23	climb stairs					1.9	1.05* (1.01,1.08)	**0.90*** **(0.86,0.95)**	**0.84*** **(0.74,0.94)**	0.97 (0.87,1.07)
24	climb 2+stairs				✓	2.6	**1.12*** **(1.08,1.16)**	**0.88*** **(0.84,0.93)**	0.96 (0.86,1.07)	**0.83*** **(0.75,0.91)**
25	lift heavy object				✓	1.5	**1.10*** **(1.07,1.14)**	0.97 (0.92,1.01)	**1.32*** **(1.13,1.48)**	**0.81*** **(0.74,0.89)**
26	move heavy object				✓	2.7	**1.13*** **(1.09,1.17)**	0.98 (0.93,1.02)	**1.41*** **(1.18,1.57)**	**0.87*** **(0.79,0.96)**
27	drive 5 miles					17.4	**1.16*** **(1.12,1.21)**	**0.73*** **(0.69,0.77)**	**2.26*** **(2.00,2.56)**	0.96 (0.86,1.07)
28	wait in line 15 min				✓	2.4	1.06* (1.02,1.09)	0.96 (0.91,1.00)	1.02 (0.91,1.16)	**0.82*** **(0.74,0.90)**
29	outside work				✓	5.4	1.06* (1.03,1.10)	1.02 (0.97,1.07)	**1.18*** **(1.05,1.32)**	0.90 (0.82,1.00)
30	Sports			✓		6.8	**1.10*** **(1.07,1.14)**	1.04 (0.99,1.09)	**0.71*** **(0.64,0.80)**	**0.77*** **(0.70,0.85)**

*false discovery rate (FDR) adjusted p-value <0.05 indicates significant DIF between sub-groups, in bold: 0.9 ≤ odds ratio (OR) ≥ 1.1 (Δ ≥ 10%) and significant FDR, SD= standard deviation, MHAQ=Modified HAQ, MD-HAQ=Multi-Dimensional HAQ, DIF=Differential Item Functioning

[1)] classified according to categories described in HAQ-DI, [2)] Items contained in the revised HAQ versions (MHAQ, MD-HAQ or HAQ-II) only

Table 3: Influence of differential item functioning on socio-demographic and disease related factors in multilevel regression models

	HAQ-DI [1)] β-coefficient (95% Confidence Interval)		**HAQ II** β-coefficient (95% Confidence Interval)		**Modified HAQ** β-coefficient (95% Confidence Interval)		**Multi-dimensional HAQ** β-coefficient (95% Confidence Interval)	
	Unadjusted	Adjusted for DIF results	unadjusted	Adjusted for DIF results	unadjusted	Adjusted for DIF results	Unadjusted	Adjusted for DIF results
	N=6176	N=6176	N=6269	N=6269	N=7058	N=7058	N=7035	N=7035
Mean HAQ scores	0.68	0.74	1.01	1.08	0.58	0.63	0.76	0.78
Age (in 10 years)	0.03* (0.02 to 0.04)	0.03* (0.02 to 0.04)	0.06* (0.04 to 0.07)	0.06* (0.05 to 0.07)	0.02* (0.01 to 0.02)	0.02* (0.01 to 0.03)	0.03* (0.02 to 0.04)	0.03* (0.02 to 0.04)
Gender (reference: male)	0.16* (0.13 to 0.19)	0.23* (0.2 to 0.26)	0.23* (0.20 to 0.26)	0.26* (0.22 to 0.29)	0.08* (0.05 to 0.11)	0.11* (0.08 to 0.14)	0.10* (0.07 to 0.13)	0.17* (0.14 to 0.19)
Education (in years at school)	-0.01* (-0.01 to -0.01)	-0.01* (-0.01 to -0.01)	-0.01* (-0.02 to -0.01)	-0.01* (-0.02 to -0.01)	-0.01* (-0.01 to -0.01)	-0.01* (-0.01 to -0.01)	-0.01* (-0.01 to -0.01)	-0.01* (-0.01 to -0.01)
Work status** (reference: not working)	-0.21* (-0.24 to -0.18)	-0.22* (-0.25 to -0.20)	-0.26* (-0.29 to -0.23)	-0.29* (-0.32 to -0.26)	-0.17* (-0.20 to -0.15)	-0.20* (-0.23 to -0.18)	-0.20* (-0.23 to -0.18)	-0.20* (-0.23 to -0.18)
DAS28*** (0-10)	0.20* (0.19 to 0.21)	0.19* (0.18 to 0.19)	0.22* (0.21 to 0.22)	0.21* (0.20 to 0.22)	0.18* (0.17 to 0.18)	0.17* (0.17 to 0.18)	0.19* (0.18 to 0.20)	0.18* (0.18 to 0.19)
RDCI** - Comorbidity Index** (0-8)	0.07* (0.06 to 0.09)	0.07* (0.06 to 0.08)	0.10* (0.08 to 0.12)	0.10* (0.08 to 0.11)	0.07* (0.05 to 0.08)	0.06* (0.05 to 0.08)	0.08* (0.07 to 0.10)	0.08* (0.07 to 0.09)
Cons	-0.23* (-0.34 to -0.13)	-0.20* (-0.30 to -0.10)	-0.18* (-0.29 to -0.07)	-0.08* (-0.19 to 0.03)	-0.14* (-0.24 to -0.05)	-0.09* (-0.18 to 0)	-0.09* (-0.19 to 0.01)	-0.09* (-0.18 to 0.00)
Country (SD cons)	0.12 (0.09 to 0.17)	0.12 (0.09 to 0.16)	0.14 (0.10 to 0.18)	0.13 (0.10 to 0.18)	0.12 (0.09 to 0.17)	0.12 (0.09 to 0.16)	0.13 (0.09 to 0.17)	0.12 (0.08 to 0.16)
Snijders/Bosker R-squared	0.45	0.47	0.47	0.48	0.41	0.43	0.45	0.46

1) HAQ-DI calculated as mean across 20 individual items, *p<0.05, HAQ= Health Assessment Questionnaire DIF=Differential item functioning **Working includes (full- and part-time employment, unpaid homemaker and students), ***DAS28=disease activity score in 28 joints, ****RDCI=Rheumatic diseases comorbidity index